

A SINGLE SERVER QUEUE WITH RANDOM ARRIVALS AND BALKING:  
CONFIDENCE INTERVAL ESTIMATION

by

Gail Rubin

and

Douglas S. Robson

BU-1019-M

April 1989

# A SINGLE SERVER QUEUE WITH RANDOM ARRIVALS AND BALKING: CONFIDENCE INTERVAL ESTIMATION\*

by

Gail Rubin and Douglas S. Robson  
Biometrics Unit, 337 Warren Hall  
Cornell University, Ithaca, NY 14853, U. S. A.

## ABSTRACT

This paper investigates a queueing system, which consists of Poisson input of customers, some of whom are lost to balking, and a single server working a shift of length  $L$  and providing a service whose duration can vary from customer to customer. If a service is in progress at the end of a shift, the server works overtime to complete the service. This process was motivated by the behavior of fishermen interviewed in the NY Great Lakes Creel Survey.

We derive the distributions of the number of services ( $X$ ), overtime and total server idle time ( $T$ ), both unconditionally (for Poisson arrivals) and conditionally on the number ( $n$ ) of arrivals per shift, assuming that the arrival times are not recorded in the data. These distributions provide the basis for estimation of the parameters from a single realization of the queueing process during  $[0, L]$ . The conditional distributions also can be used to estimate common service time,  $w$ , when  $(n, X)$  or  $(n, T)$  are observed. Confidence intervals based on  $T$  are of shorter length, for all confidence coefficients, than the corresponding intervals based on  $X$ .

Keywords: Single server queue, balking, conditional distributions, maximum likelihood, confidence limits.

Short Title: Single Server Queue With Balking

\* This paper is Technical Report #BU-1019-M in the Biometrics Unit Series. The authors are grateful to N.U. Prabhu for suggestions on streamlining the distributional derivations and to D.R. Cox and C.E. McCulloch for helpful comments.

## 1. INTRODUCTION

The queueing system to be considered consists of Poisson input of customers, some of whom are lost to balking, and a single server working a shift of length  $L$  and providing a service whose duration,  $w_i$ , can vary from customer to customer. Investigation of this queueing system was motivated by the behavior of fishermen encountered in the NY Great Lakes Creel Survey (Robson and Jones [9]). In this study, an interviewer is stationed at a marina or boat launch for a predetermined shift length and asks fishermen returning with their catch a fixed set of questions, requiring constant service time,  $w$ . Balking arises since a fisherman will leave immediately if the interviewer is occupied. Although the model described here will allow for variable service time, the length of the interview was virtually constant for all fishermen. If a service is in progress at the end of a shift, the server works overtime to complete the service. Consequently, no queue accumulates: the customer's waiting time is always zero and the customer's time in the system is either 0 or  $w_i$ . Feller [4, pp. 306, 315, 339] describes a process for the Type I Geiger counter, which is similar but lacks the overtime feature, and gives the asymptotic mean and variance of the number of particle registrations, i.e., the number of services. For the Creel Survey, the goal is to estimate the unknown number ( $n$ ) of fishermen returning to the marina during the shift, based on the known number of interviews. Once the number of arrivals is estimated, the number of fish caught can be estimated for the balkers, thereby providing the estimates of the total catch for each fish species.

Several features of the Creel Survey make application of this simple queueing model appropriate. The interviewer visits several marinas in a work day, starting at a random location along the route. The shift lengths for each marina on any day are fixed and range from one to two hours. Because (1) the shift length at any site is short relative to the length of a fishing day and (2) the time at which the interviewer arrives at a given marina is random (due to the random starting point in his route), the assumption of Poisson arrival process of fisherman seems reasonable. Recall that the purpose of applying the queueing model to the Creel Survey data is to allow estimation of the number of arrivals for each shift at each marina, which will allow estimation of the number of fish caught by the balkers. The data base contains information from several shifts at each marina for a given interviewer. The arrival rate parameter will vary from site to site and from day to day for any site due to weather conditions, time of day that the interviewer arrives at the site, etc. Thus, to get the best correction for missing values (i.e., the number of fish caught by the balkers), one wants to estimate the number of arrivals separately for each shift and each site.

The distributional properties of the number of services performed ( $X$ ), overtime ( $Y$ ) and total server idle time ( $T$ ) are derived both unconditionally (for Poisson arrivals) and conditionally on the number of arrivals per shift in Sections 2 and 3, assuming that arrival times are not recorded in the data. We consider the distributional results for overtime more appropriate for measuring queueing system behavior rather than for estimating the number of arrivals ( $n$ ) or the arrival rate ( $\lambda$ ). The service times are assumed to be independent,

identically distributed, strictly positive random variables. However, for purposes of estimation, it is sometimes useful to view the service time distribution in terms of a given sequence of  $w_i$ 's that is known through the final service of the shift.

The statistical estimation discussed here and in Rubin and Robson [11] is based on observation of a single shift from the queueing process, which often is the case in nonindustrial applications. This differs somewhat from estimation and statistical inference for stochastic processes discussed by Basawa and Prakasa Rao [1] and Basawa and Prabhu [2] in that, with a single observation of the process, one cannot rely on large sample properties. We do not consider the consistency of the estimators given here; our concern focuses primarily on exact (small sample size) properties.

Samaan and Tracy [12] derive estimators of the customer arrival rate for a single server queue with loss (balking) when the customer arrivals form a Poisson process and the service times are uniform on the interval  $(0, 1)$ . In their case, the interdeparture times of customers are known and become the basis for estimation of  $\lambda$ . Basawa and Prabhu [2] discuss large sample inference for single server queues, having interarrival time and service time distributions belonging to the exponential family, where the sample data are assumed to contain the set of interarrival times and service times as well as the number of arrivals and the number of service completions. They derive approximate maximum likelihood estimates of the customer arrival rate for several models. The stopping rule for the process described above is a hybrid of rules 1 and 2 of Basawa and Prabhu [2] due to the overtime feature. However, when one uses

the distributions that are conditional on the number of arrivals per shift for estimation, one operates under a rule, which is a hybrid of Basawa and Prabhu's rules 1, 2 and 3, which trivially satisfies the conditions for their approximate likelihood function. However, the stability conditions required for application of their estimators cannot be met with a single observation of the process.

In this paper, point and interval estimators of the unknown number of arrivals ( $n$ ) are derived from the conditional distributions of total server idle time ( $T$ ) or the number of services ( $X$ ), which are given in Section 3. Derivation of corresponding estimators of the unknown rate ( $\lambda$ ) of the Poisson arrival process, based on the unconditional distributions, are described briefly. We also consider estimation of common service time ( $w$ ) for the case of service times assumed to be equal, when either ( $n, X$ ) or ( $n, T$ ) are observed from a single shift of the queueing process. Statistical estimates of parameters, such as service time, from a system in operation aid in assessing queueing system behavior and in design of more efficient systems (Prabhu [8]).

For notational convenience, variables, moments, probabilities and distributions, which are conditional on the realized number ( $n$ ) of arrivals, will be denoted with a lowercase subscript ( $n$ ), while their unconditional counterparts bear an uppercase subscript ( $N$ ). We use the term probability density function (pdf) loosely, applying it to mixed distributions as well as to continuous distributions. The special case of constant service time,  $w$ , for all customers will be discussed where this simpler case affords results unavailable in general.

## 2. JOINT DISTRIBUTION OF NUMBER OF SERVICES AND OVERTIME

The joint pdf of the number of services and overtime can be used as the basis for deriving all other distributional results. Consequently, we will derive it for the general case of unequal service times.

Assume that the sequence of service times  $w_1, w_2, \dots$  is known through the final service of the shift or that these observed  $w_i$  are the given realization of independent, identically distributed (strictly positive) random variables. Let  $W_i$  denote the partial sums of this sequence,  $W_0 \equiv 0$  and  $W_i = W_{i-1} + w_i$ . Let  $Z_i$  denote the number of balkers arriving during the  $i^{\text{th}}$  service and let  $\eta_i$  denote the partial sums,  $\eta_0 \equiv 0$  and  $\eta_i = \eta_{i-1} + Z_i$ . If  $N$  customers arrive during  $(0, L)$  and exactly  $X$  services are performed then  $\eta_X = N - X$ . Note that the random variable  $W_{X-1}$  is bounded above by  $L$ ;  $I_{(0, L]}(L - W_{X-1}) = 1$ . Hence, for a given  $w_i$  sequence,  $X$  is bounded above by  $M$  such that  $W_{M-1} \leq L < W_M$ , and when  $n$  also is given then  $X \leq \min(n, M)$ .

Let  $I_j$  denote the duration of the server's idle period immediately preceding the start of the  $j^{\text{th}}$  service. Thus,  $I_1$  is the arrival time of the first of the  $N$  customers and for  $j > 1$ ,  $I_j$  is the length of time between completion of the  $(j-1)^{\text{th}}$  service and the arrival of the next customer. We note that the Poisson input assumption implies that  $I_1, I_2, \dots$  is a sequence of independent, identically distributed exponential random variables.

Let  $T$  denote the server's total idle time,  $0 < T < L$ , and define

$$T = L - W_X + Y = L - W_{X-1} - (w_X - Y) \quad (1)$$

where  $Y$  then denotes the amount of overtime,  $0 \leq Y < w_x$ . Note that when  $L$  and the service time sequence  $w_1, w_2, \dots$  are specified, then specification of  $T$  uniquely determines  $X$  and  $Y$ . For the case of all service times equal, when  $L$  and the common service time,  $w$ , are known, one can solve explicitly for  $X$  from  $T$  without knowing the service time sequence, using the relation:

$$X-1 < (L-T) / w \leq X.$$

This joint distribution is obtainable by formally invoking the Poisson assumption to give, for  $0 < y < w_x$ ,  $P(X=x, Y>y)$

$$\begin{aligned} &= P(l_1+w_1+l_2+w_2+\dots+l_x \leq L < L+y < l_1+w_1+l_2+w_2+\dots+l_x+w_x) \\ &= P(l_1+l_2+\dots+l_x \leq L-W_{x-1} < L-W_{x-1}+y < l_1+l_2+\dots+l_x+w_x) \\ &= P(L-W_x+y < l_1+l_2+\dots+l_x \leq L-W_{x-1}) \\ &= G_x(L-W_{x-1}; \lambda) - G_x(L-W_x+y; \lambda) \end{aligned}$$

or, for  $L-W_x < t \leq L-W_{x-1}$ ,

$$P(X=x, T > t) = G_x(L-W_x; \lambda) - G_{x+1}(t; \lambda),$$

where the gamma cdf  $G_x(t; \lambda)$  is the  $x$ -fold convolution of  $1-\exp(-\lambda t)$ .

Similarly, at  $y = 0$ , we formally obtain a Poisson frequency function for

$$\begin{aligned} P(X=x, Y=0) &= P(l_1+l_2+\dots+l_x \leq L-W_x < l_1+l_2+\dots+l_{x+1}) \\ &= G_x(L-W_x; \lambda) - G_{x+1}(L-W_x; \lambda) \\ &= P(X=x, T=L-W_x). \end{aligned}$$

Differentiating produces the following:

**Theorem 1.** The unconditional joint density function of the number of services,  $X_N$ , and overtime,  $Y_N$ , when the customer arrivals form a Poisson process with rate parameter  $\lambda > 0$ , is:



$$f_{X_N, Y_N}(x, y; \lambda, L) =$$

$$\begin{cases} \exp(-\lambda(L-W_x)) (\lambda(L-W_x))^x / x! & \text{for } y=0, W_x < L \\ \lambda \exp(-\lambda(L-W_x+y)) (\lambda(L-W_x+y))^{x-1} / (x-1)! & \text{for } \max(0, W_x - L) < y \leq W_x \\ 0 & \text{otherwise.} \end{cases}$$

Conditional on  $X$  and  $T$ , the frequency distribution of the number of balkers arriving during the fixed interval  $(0, L)$ ,  $\eta_x = N - X$ , is the Poisson distribution of the number of arrivals during a period of total duration  $L - T$ . Hence, for

$$0 \leq X \leq \min(n, M - 1) \text{ and } n \geq 0,$$

$$\begin{aligned} f_{N, X_N, T_N}(n, x, L - W_x; \lambda, L) \\ = \exp(-\lambda(L - W_x)) (\lambda(L - W_x))^x \exp(-\lambda W_x) (\lambda W_x)^{n-x} / (x! (n-x)!) \end{aligned}$$

and, for  $L - W_x < t \leq L - W_{x-1}$ ,  $1 \leq X \leq \min(n, M)$  and  $n \geq 1$ ,

$$\begin{aligned} f_{N, X_N, T_N}(n, x, t; \lambda, L) \\ = \lambda^x t^{x-1} \exp(-\lambda t) (\lambda(L - t))^{n-x} \exp(-\lambda(L - t)) / ((x-1)! (n-x)!). \end{aligned}$$

Since  $N$  is Poisson distributed with parameter  $\lambda L$ , we obtain the conditional distributions of  $(X_n, Y_n)$  or  $(X_n, T_n)$  as

**Theorem 2.** The joint density function of the number of services,  $X_n$ , and overtime,  $Y_n$ , conditional on  $N = n$  arrivals, where  $n > 0$ , is:

$$f_{X_n, Y_n}(x, y; n, L) = \begin{cases} \binom{n}{x} (L - W_x)^x W_x^{n-x} / L^n & \text{for } W_x < L, y = 0 \\ x \binom{n}{x} (L - W_x + y)^{x-1} (W_x - y)^{n-x} / L^n & \text{for } \max(0, W_x - L) < y < W_x \\ 0 & \text{otherwise,} \end{cases}$$

where  $X \leq \min(n, M)$ .

Exactly the same distributions for  $(X, Y)$  or  $(X, T)$  would be obtained under an alternative scenario in which service is instantaneous but the server's work shift

is foreshortened by an amount  $w_i$  with the arrival of the  $i^{\text{th}}$  customer, subject only to the overtime constraint; i.e., the server gives up an amount of time off equal to  $y$ . We refer to this alternative scenario as the "instant service mode" in contrast to the earlier "real time mode". Probability calculations are sometimes more easily seen in one mode than the other (see Appendix B).

### 3. MARGINAL DISTRIBUTIONS OF TOTAL IDLE TIME, NUMBER OF SERVICES AND OVERTIME

The unconditional marginal distributions of  $T$ ,  $X$ , and  $Y$  can be derived as corollaries to Theorem 1, while their conditional counterparts can be derived as corollaries to Theorem 2. As usual, the marginal results for a given variable are derived by integrating over or summing out the other variable in the bivariate distributions.

#### 3(a) Total idle time ( $T$ )

Notice that when the sequence of service times is known through the final service,  $T$  uniquely specifies  $X$ . Consequently, the pdf of idle time is given by:

Corollary 1.1. For a given sequence of service times, the pdf of total server idle time ( $T_N$ ), when customer arrivals form a Poisson process with rate  $\lambda > 0$ , is:

$$f_{T_N}(t; \lambda, L) = \begin{cases} (\lambda t)^x \exp(-\lambda t) / x! & \text{for } t = L - W_x \\ \lambda (\lambda t)^{x-1} \exp(-\lambda t) / (x-1)! & \text{for } \max(0, L - W_x) < t < L - W_{x-1} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The cdf and upper tail probability can be derived directly from the pdf. The cdf has jump discontinuities corresponding to increases in the number of services.

Rubin [10] discusses a generalized unconditional cdf of idle time for equal service times, which covers the scenario of a time homogeneous Poisson arrival process with parameter  $\lambda$ , but  $\lambda$  itself is the realization of a random variable with cdf  $H$ . Consequently, each of multiple observations of the process could have a different rate parameter for the Poisson arrival process, and with idle time observed for each replicate, the mixing distribution  $H$  becomes the estimation target.

The marginal distribution of total server idle time, conditional on the number of arrivals ( $n$ ), can be derived from Theorem 2.

Corollary 2.1. The pdf of total server idle time ( $T_n$ ), conditional on the number of arrivals ( $n$ ) during a shift of length  $L$ , is given by:

$$f_{T_n}(t; n, L) = \begin{cases} \binom{n}{x} t^x (L-t)^{n-x} / L^n & \text{for } t = L - W_x \\ x \binom{n}{x} t^{x-1} (L-t)^{n-x} / L^n & \text{for } \max(0, L - W_x) < t < L - W_{x-1} \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where  $\max(0, L - W_n) \leq t < L$ ,  $0 < w_x$ .

The conditional cdf and conditional upper tail probability can be derived directly from the pdf; they bear superficial resemblance to an incomplete beta function, but the arguments depend on  $t$ . For  $0 < t < L$ , the cdf is strictly increasing unless  $0 \leq t < L - W_n$ , when  $F_{T_n}(t) = 0$ , and the cdf jumps at  $t = L - W_j$ , for  $j = 1, \dots, n$ . The cdf approaches the limit one as  $t$  approaches  $L$ , but it is actually undefined at  $t = L$ .

Figure 1 illustrates the conditional cdf for equal service times with  $L=1$ ,  $w=0.1$  and  $n = 20$ , as well as the unconditional cdf with  $L=1$ ,  $w = 0.1$  and

Poisson rate parameter,  $\lambda=20$ . The conditional and unconditional cdfs of idle time can be used to construct interval estimators for  $n$  and  $\lambda$ , respectively, while the pdfs of idle time allow for maximum likelihood estimation of  $n$  and  $\lambda$ .

The unconditional and conditional expected values of total server idle time do not have a compact form. For the case of equal service time, they are most easily calculated using the linearity property of expectations and substituting the appropriate expectations of  $X$  and  $Y$ .

### 3(b) Number of services (X)

The marginal distributions of the number of services ( $X$ ) become relevant when the amount of overtime ( $Y$ ) is unobservable. The unconditional marginal distribution of the number of services can be written most succinctly as the upper tail probability, which is the following cumulative chi square probability.

Corollary 1.2. The unconditional upper tail probability of the number of services,  $X_N$ , is:

$$\begin{aligned} P(X_N \geq x; \lambda, L) &= \int_0^{\lambda(L-W_{x-1})} q^{x-1} \exp(-q) dq / \Gamma(x) \\ &= P\left(\chi_{2x}^2 < 2\lambda(L-W_{x-1})\right) \end{aligned} \quad (4)$$

for  $1 \leq x \leq M$  and  $\lambda > 0$ .

Note that  $P(X_N \geq 0; \lambda, L) = 1$ . The expected value for the unconditional distribution,

$$E(X_N) = \sum_{x=1}^M \int_0^{(L-W_{x-1})} \lambda^x q^{x-1} \exp(-\lambda q) dq / \Gamma(x) \quad (5)$$

for  $0 \leq W_{x-1} < L$  and  $M: W_{M-1} \leq L < W_M$ , does not have a simple closed form.

For equal service time, the unconditional expected number of services can be approximated as:

$$E(X_N) \cong \lambda L / (1 + \lambda w) + 0.5 (\lambda w / (1 + \lambda w))^2, \quad (6)$$

for fixed  $\lambda w$  and large  $\lambda$ . This approximation results from an argument of approximate stationarity of the queueing process for large  $\lambda$  and small  $w$ .

An intuitive rationale for (6) is that the average number of services should be (average number of arrivals) / (average number of arrivals per service), where the denominator is  $1 + \lambda w$ , the served customer plus the expected number of balkers arriving during the ensuing busy period. If overtime occurs, then arrivals during the overtime period should be included in the numerator, since the denominator,  $1 + \lambda w$ , treats all service periods as complete. Thus, the numerator is the average number of arrivals in an interval  $(0, L + Y)$  of random length  $L + Y$ . Conditional on  $Y_N$ , the expected number of arrivals is  $\lambda(L + Y_N)$ . Therefore, the ratio above becomes:

$$\lambda (L + E(Y_N)) / (1 + \lambda w). \quad (7)$$

As will be shown in Section 3(d), the unconditional expected value of overtime can be approximated by:

$$E(Y_N) \cong \lambda w^2 / (2 (1 + \lambda w)). \quad (8)$$

Substituting (8) into (7) gives (6).

Cox and Smith [3, pp. 49, 65-68] note that stationary processes often provide good approximations to similar processes lacking stationarity. The approximation given in (6), like those which will be presented for overtime, can be shown numerically to behave very well for small to moderate  $w$  and small to

large  $\lambda$  (Table 1).

Table 1. The exact and approximate unconditional expected number of services,  $E(X_N)$ , for several values of  $\lambda$  and  $w$  with  $\lambda w$  held constant and  $L=1$ . The % relative error of the approximation has been calculated as  $((\text{exact} - \text{approximate}) / \text{exact}) \times 100$ .

$\lambda$	$w$	$\lambda w$	$E(X_N)$		% Relative Error
			Exact	Approximate	
2.0	0.5000	1.0	1.12890583	1.12500000	0.34598
3.0	0.3333	1.0	1.62450848	1.62500000	-0.03026
4.0	0.2500	1.0	2.12484783	2.12500000	-0.00716
5.0	0.2000	1.0	2.62503016	2.62500000	0.00115
6.0	0.1667	1.0	3.12500570	3.12500000	0.00018
7.0	0.1429	1.0	3.62499832	3.62500000	-0.00005
8.0	0.1250	1.0	4.12499832	4.12500000	-0.00000
9.0	0.1111	1.0	4.62500009	4.62500000	0.00000
10.0	0.1000	1.0	5.12500000	5.12500000	0
25.0	0.0400	1.0	12.62500000	12.62500000	0
50.0	0.0200	1.0	25.12500000	25.12500000	0
100.0	0.0100	1.0	50.12500000	50.12500000	0
2.0	0.2500	0.5	1.38889234	1.38888889	0.00025
3.0	0.1667	0.5	2.05555549	2.05555556	0.00000
4.0	0.1250	0.5	2.72222222	2.72222222	0
5.0	0.1000	0.5	3.38888889	3.38888889	0
6.0	0.0833	0.5	4.05555556	4.05555556	0
7.0	0.0714	0.5	4.72222222	4.72222222	0
8.0	0.0625	0.5	5.38888889	5.38888889	0
9.0	0.0556	0.5	6.05555556	6.05555556	0
10.0	0.0500	0.5	6.72222222	6.72222222	0
25.0	0.0200	0.5	16.72222222	16.72222222	0
50.0	0.0100	0.5	33.38888889	33.38888889	0
100.0	0.0050	0.5	66.72222222	66.72222222	0

As for the unconditional marginal distribution, the marginal distribution of the number of services ( $X_n$ ), conditional on the number of arrivals ( $n$ ), can be written most succinctly as the upper tail probability. Notice that the conditional upper tail probability corresponds to a cumulative beta probability.

Corollary 2.2. The upper tail probability of the number of services,  $X_n$ , conditional on the number of arrivals, is:

$$P(X_n \geq x; L) = x \binom{n}{x} \int_0^{1-(W_{x-1}/L)} u^{x-1} (1-u)^{n-x} du \quad (9)$$

$$= \sum_{r=0}^{n-x} \binom{n}{r} W_{x-1}^r (L-W_{x-1})^{n-r} / L^n \quad (10)$$

for  $0 \leq W_{x-1} < L$  and  $x \leq \min(n, M)$ .

Recalling the connection between an incomplete beta integral and a truncated binomial sum provides the sum given in (10) (Johnson and Kotz [6]). The expected value for the conditional distribution,

$$E(X_n) = \sum_{x=1}^{\min(n, M)} I_{(0, L]}(L-W_{x-1}) \sum_{r=0}^{n-x} \binom{n}{r} W_{x-1}^r (L-W_{x-1})^{n-r} / L^n, \quad (11)$$

also lacks a simple closed form.

For the special case of equal service time for all customers, approximate stationarity of the process occurs for large arrival rate ( $\lambda$ ) and short service times ( $w$ ), suggesting the approximation:

$$E(X_n) \cong n / (1 + (n-1)(w/L)), \quad (12)$$

which is exact for  $w = 0$ ,  $w = L/(n-1)$  and  $w = L$ . Notice that the denominator of the approximation is the expected number of clients that balk or are served per

service time  $w$ , while the numerator is the total number of arrivals in the shift. The approximation behaves well for a variety of  $n$ ,  $L$ ,  $w$  configurations (Figure 2). Moreover, for the case of equal service time, the conditional cdf of the number of services ( $X_n$ ) and the conditional upper tail probability of the number of balkers ( $n - X_n \equiv Z_n$ ) are equivalent (Rubin [10]). This allows one to derive estimators of  $n$  when only the number of balkers is observed.

$T$  is stochastically larger than  $X$ , both conditionally and unconditionally, since either pair of distributions for the general case has monotone likelihood ratio (Lehmann [7], pp. 83-85). The monotonicity of the ratio of beta kernels of  $X$  and  $T$  (conditional) or the ratio of gamma kernels (unconditional), and the fact that, for either set of kernels, the integral of the kernel of  $X$  over the entire range equals that of  $T$ , imply that the cutoff for  $X$  is less than the cutoff for  $T$  for all  $\alpha \in (0, 1)$ . Consequently, for any parameter, the  $1-2\alpha$  confidence interval based on  $T$  is shorter than corresponding interval based on  $X$  for all  $\alpha \in (0, 1)$ . This is satisfying from a statistical point of view: the interval estimator based on more information ( $T$ ) is more efficient, in the sense of having shorter length.

### **3(c) Overtime ( $Y$ )**

The marginal distributions of overtime and approximations to relevant quantities are given in this section. The approximations for the probability of no overtime and for the mean and variance of overtime are useful in measuring the behavior of queueing systems. Simplified conditional results exist for the case of equal service times ( $w$ ) with  $n < [L/w]$ , where  $[ \cdot ]$  designates the integer part of the argument; they are given in Section 3(d).



The unconditional distribution of overtime is derived directly from the unconditional joint density function of the number of services and overtime given in Theorem 1.

Corollary 1.3. The unconditional pdf of overtime,  $Y_N$ , when customer arrivals form a Poisson process with rate  $\lambda > 0$ , is:

$$f_{Y_N}(y; \lambda, L) = \begin{cases} \sum_{x=0}^{M-1} \exp(-\lambda(L-W_x)) (\lambda(L-W_x))^x / x! & \text{for } y = 0 \\ \sum_{x=1}^M \lambda I_{(\max(0, W_x-L), w_x)}(y) \exp(-\lambda(L-W_x+y)) (\lambda(L-W_x+y))^{x-1} / (x-1)! & \text{for } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where  $M: W_{M-1} \leq L < W_M$ .

The exact unconditional mean and variance of overtime are quite messy, even for equal service times (Rubin [10]). However, in the case of equal service time, many quantities calculated for the unconditional distribution of overtime can be approximated making use of the approximate stationarity of the process. The following approximations for the unconditional distribution of overtime have been shown empirically to be well-behaved for fixed  $\lambda w$  and large  $\lambda$ :

$$P(Y_N = 0; \lambda, w) \cong 1 / (1 + \lambda w) \quad (13)$$

$$f_{Y_N}(y; \lambda, w) \cong \lambda w / (1 + \lambda w) \quad (14)$$

$$F_{Y_N}(y; \lambda, w) \cong (1 + \lambda y) / (1 + \lambda w) \quad (15)$$

$$E(Y_N) \cong \lambda w^2 / (2(1 + \lambda w)) \quad (16)$$

$$\text{Var}(Y_N) \cong \lambda w^3 (4 + \lambda w) / (12(1 + \lambda w)^2) \quad (17)$$

The rationale for these approximations is as follows. Recall that a shift

consists of pairs of busy and idle times for the server, and the occurrence of overtime depends on the random positioning of a shift termination point within the busy/idle pair. Thus, the probability of positive overtime is equivalent to the probability of the shift termination point occurring during a busy period (the constant service time,  $w$ ). Applying the renewal theorem to the limiting case of the queueing process gives:

$$\begin{aligned} P(\text{positive overtime}) &\cong \frac{E(\text{busy period})}{E(\text{busy period}) + E(\text{idle period})} \\ &= w / (w + \lambda^{-1}) = \lambda w / (1 + \lambda w), \end{aligned}$$

which equals (14). Consequently, the probability of no overtime, for the limiting case, is:

$$\begin{aligned} P(Y_N = 0; \lambda, w) &= 1 - P(\text{positive overtime}) \\ &\cong 1 / (1 + \lambda w), \end{aligned}$$

which equals (13).

Notice that the random positioning of a termination point implies that if the point does fall in a busy period, then overtime is uniformly distributed on the interval  $(0, w)$ . This observation, together with (13) and (14) give the cdf of overtime for the stationary process as:

$$\begin{aligned} F_{Y_N}(y; \lambda, w) &\cong 1 / (1 + \lambda w) + (\lambda w / (1 + \lambda w)) \int_0^y (1 / w) dt \\ &= (1 + \lambda y) / (1 + \lambda w), \end{aligned}$$

which equals (15). Similarly, the expected value of overtime for the stationary process is:

$$E(Y_N) \cong (1+\lambda w)^{-1}(0) + (\lambda w / (1+\lambda w)) w^{-1} \int_0^w y \, dy ,$$

while the second moment of  $Y_N$  is approximately  $\lambda w^3 / (3(1+\lambda w))$ .

Consequently, the variance of overtime can be approximated by (17).  $\square$

Numerical results for  $P(Y_N=0)$ ,  $E(Y_N)$  and  $\text{Var}(Y_N)$  indicate that the approximations are good to at least eleven decimal places for  $w < 0.2$  and moderate to large  $\lambda$  ( $\lambda > 6$ ) and are virtually exact for  $0.001 < w < 0.1$  and  $\lambda > 1$ . The approximations are good to at least six decimal places for  $0.2 < w < 0.4$  and moderate to large  $\lambda$  ( $\lambda > 10$ ).

The conditional marginal distribution follows directly from the conditional joint distribution of the number of services and overtime given in Theorem 2.

**Corollary 2.3.** The pdf of overtime,  $Y_n$ , conditional on the number of arrivals, is:

$$f_{Y_n}(y; n, L) = \begin{cases} \sum_{x=1}^{\min(n, M)} \binom{n}{x} (L - W_x)^x W_x^{n-x} / L^n & \text{for } y = 0 \\ \sum_{x=1}^{\min(n, M)} I_{(\max(0, W_x - L), w_x)}(y) \times \binom{n}{x} (L - W_x + y)^{x-1} (W_x - y)^{n-x} / L^n & \text{for } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

for  $M$ :  $W_{M-1} \leq L < W_M$ .

The conditional cdf and conditional expectation can be found by integration, but the latter does not have a compact form.

### 3(d) Conditional distribution of overtime for the case of $n < [L/w]$

The conditional distribution of overtime simplifies greatly for the special case of equal service time for all customers and the number of arrivals less than  $[L/w]$ , the maximum number of services possible in a shift of length  $L$ . Recursion formulae exist for the  $P(Y_n = 0)$  and the  $P(Y_n > y)$  in this special case. These are helpful in performing numerical and algebraic calculations. For typographic simplicity, we let the shift length equal unity ( $L=1$ ) in this section.

Corollary 2.4. When all services are of duration  $w$  and  $L=1$ , the probability of no overtime, conditional on the number of arrivals, is:

$$P(Y_n = 0) = 1 - nw P(Y_{n-1} = 0) \quad \text{for } n < [1/w].$$

The recursion relation is proved by showing that

$$P(Y_n = 0) + nw P(Y_{n-1} = 0) = 1.$$

This requires several binomial expansions in conjunction with recombining of terms, which allows one to rewrite the resulting expression as a polynomial in  $w$ . Details of the proof are given in Appendix A.

Corollary 2.5. When all services are of duration  $w$  and  $L=1$ , the probability of no overtime, conditional on the number of arrivals is

$$P(Y_n = 0) = \sum_{i=1}^n (n)_i (-w)^i,$$

where  $(n)_i = \prod_{k=1}^i (n-k+1)$  for  $n < [1/w]$ .

Proof. From Corollary 2.4 we know that

$$P(Y_n = 0) = 1 - nw \sum_{x=0}^{n-1} \binom{n-1}{x} (1-xw)^x (xw)^{n-x}. \quad (18)$$

Expanding the finite sum and rearranging terms give the above formula.  $\square$

Corollary 2.6. When all services are of duration  $w$  and  $L=1$ , the probability that overtime exceeds  $y$ , conditional on the number of arrivals, is:

$$P(Y_n > y) = n(w-y) P(Y_{n-1} = 0) \quad (19)$$

for  $n < [1/w]$  and  $0 < y < w < 1$ .

A sketch of the proof is given in Appendix B.

For  $n < [1/w]$ ,  $P(Y_n = 0)$  can be approximated by

$$1/(1+(n+0.5)w). \quad (20)$$

The core of the approximation,  $1/(1+nw)$ , results from the approximate stationarity argument for the queueing process. Empirical results indicate that the added  $1/(2w)$  in the denominator improves the approximation. An approximation based on the same reasoning is available for  $P(Y_n > y)$  with  $n < [1/w]$ :

$$P(Y_n > y) \cong n(w-y)/(1+(n-0.5)w). \quad (21)$$

As a consequence of Corollaries 2.4 and 2.6 we note that:

Corollary 2.7. When all services are of duration  $w$  and  $L=1$ , the distribution of overtime, conditional on the number of arrivals and on overtime positive, is uniform on the interval  $(0, w)$  when  $n < [1/w]$ .

Proof. Because we have Poisson arrival of customers, the arrival times are uniformly distributed on the interval  $(0, 1)$  for shift length equal to one. Positive overtime is generated by the arrival time of the last served customer occurring after  $1-w$  but before 1. By conditioning on positive overtime, we are rescaling

an arrival time distribution that is uniform on  $(0, 1)$  to be uniform on a shorter interval,  $(0, w)$ .  $\square$

The expected value and variance for this overtime distribution are  $w/2$  and  $w^2/12$ , respectively. Applying the results of Corollaries 2.4 and 2.6 gives  $P(Y_n > y | Y_n > 0)$  equal to  $(w-y)/w$ . Thus, for  $n < [1/w]$

$$F_{Y_n}(y | Y_n > 0) = y / w.$$

#### 4. ROLE OF DISTRIBUTIONAL RESULTS IN ESTIMATION

Point and interval estimators of the unknown number of arrivals ( $n$ ) or the unknown rate ( $\lambda$ ) of the Poisson arrival process can be derived from the conditional and unconditional distributions, respectively, of total server idle time ( $T$ ) or the number of services ( $X$ ). The target of inference and which distributions are relevant may differ in accordance with the type of data that have been collected. For the Creel Survey, total server idle time is unlikely to be the data. However, in an industrial setting, total server idle time or its complement, the cumulative service time for the shift, may be the relevant data.

Confidence limits for  $n$  and  $\lambda$  are derived by writing confidence statements in terms of the appropriate cdfs or upper tail probabilities and solving the resulting equations for their parameters, while maximum likelihood estimators of  $n$  and  $\lambda$  are derived using the pdfs of  $T$  or  $X$ . For the case of equal service time, Rubin and Robson [11] have constructed point estimators of  $n$ , based on  $T$  or  $X$ , that are unbiased over the restricted range of  $n < 1/w$ . They [11] also derive point and interval estimators of  $n$  using the conditional cdf of the number of balkers

(Z), for the case of equal service time. Estimators of  $n$  and  $\lambda$  could be derived from the marginal distributions of overtime, although only observing overtime seems unlikely for most applications.

In the Sections 5 and 6, the conditional distributions are used to estimate the number of arrivals ( $n$ ) or common service time ( $w$ ) when  $n$  and  $X$  or  $n$  and  $T$  are observed. Both maximum likelihood estimates and confidence limits are derived.

## **5. ESTIMATION OF NUMBER OF ARRIVALS ( $n$ )**

Interval or point estimation of  $n$  can be based on the conditional distributions of  $X$ ,  $(X, Y)$  or  $T$  when the shift length is known. One must know the sequence of  $w_i$ 's through  $x$  to use  $T$ , while one need know only the cumulative service time,  $W_x$ , to use  $(X, Y)$  or  $X$ . When the sequence of service times is known through the  $x^{\text{th}}$  service, then  $T=L-W_x+y$  uniquely specifies  $X$ . It is not possible, however, to resurrect  $(X, Y)$  from  $T$  when only  $W_x$  is known.

### **5(a) Point Estimators of $n$ Based on $T$ , $(X, Y)$ or $X$**

A maximum likelihood estimator (MLE) of the number of arrivals can be constructed either from the conditional density function of  $T$ , given in (3) or from the conditional joint density function of  $X$  and  $Y$ , given in Theorem 2. Notice that one must know the sequence of  $w_i$ 's through  $x$  to use the former, while one need know only the cumulative service time,  $W_x$ , to use the latter. In either case, setting the difference between the likelihood functions at  $n$  and  $n-1$  equal to zero yields:

$$\hat{n}_{ml}(T) \equiv x L / T = X L / (L - W_x + Y). \quad (22)$$

The ratio of the likelihoods,

$$f_n(t_x) / f_{n-1}(t_x) = (n / (n-x)) ((L - t_x) / L) = (n / (n-x)) ((W_x - y) / L),$$

where  $0 < L - t_x = W_x - y < L$ , is a strictly decreasing function of  $n$ , which passes through unity at a point  $n = \hat{n}_{ml}(T)$  that is relatively close to  $x$  for small  $W_x$ . The ratio of adjacent ratios of the likelihoods is less than unity:

$$(f_{n+1}(t_x) / f_n(t_x)) / (f_n(t_x) / f_{n-1}(t_x)) = 1 - x / (n(n-x+1)) < 1$$

for  $0 \leq x \leq n$ , implying that the likelihood function is unimodal with its maximum near  $\hat{n}_{ml}(T) = X L / (L - W_x + Y)$ .

The MLE of  $n$  based on  $X$ ,  $\hat{n}_{ml}(X)$ , does not have a closed form and is most easily found by calculating  $P(X_n = x; n, L) =$

$$\sum_{r=0}^{n-x} \binom{n}{x} W_{x-1}^r (L - W_{x-1})^{n-r} / L^n - \sum_{r=0}^{n-x-1} \binom{n}{x+1} W_x^r (L - W_x)^{n-r} / L^n$$

for successive values of  $n$  until the probability decreases. The value of  $n$ , for which  $P(X_n = x; n, L)$  is largest, is  $\hat{n}_{ml}(X)$ .

The finiteness of  $\hat{n}_{ml}(X)$  is guaranteed because it is bounded above and below by the largest and smallest values, respectively, of

$$\hat{n}_{ml}(T) = \hat{n}_{ml}(X, Y) = X L / (L - W_x + Y)$$

as a function of  $Y$ ; thus, the following theorem proves that  $P(X_n = x; n, L)$  is maximum for  $n$  in the given interval. Using the fact that  $f_n(x, y) / f_{n-1}(x, y)$  is a decreasing function of  $n$ , which passes through unity at  $n = X L / (L - W_x + Y)$  for  $0 \leq y < w_x$ , we can prove by contradiction that the  $n$ -solution to  $f_n(x) = f_{n-1}(x)$  belongs to the interval given in the theorem below.



**Theorem 3.** The integer-valued MLE of  $n$  based on  $X$ ,  $\hat{n}_{ml}(X)$ , satisfies

$$XL / (L - W_{x-1}) \leq \hat{n}_{ml}(X) \leq XL / (L - W_x)$$

for  $W_x < L$  and  $0 \leq y < w_x$ .

Maximum likelihood estimation of  $\lambda$  based on  $T$  or  $X$  is simpler than that for  $n$ , since the respective unconditional likelihood functions can be differentiated with respect to  $\lambda$ . Rubin and Robson [11] give details.

### 5(b) Confidence Limits for $n$

The cdf of idle time, conditional on  $n$  arrivals, is an increasing function of  $n$  and can be used to construct confidence limits for  $n$ . A  $1-\alpha$  lower confidence limit for  $n$  can be constructed by solving for  $n$  in the equation:

$$\alpha = P(T_n \leq t / L; n) = x^* \binom{n}{x^*} \int_0^{t/L} u^{x^*-1} (1-u)^{n-x^*} du, \quad (23)$$

where  $0 < t/L < 1$ ,  $0 < w_x$  and

$$x^* = \begin{cases} x+1 & \text{for } L-t = W_x \\ x & \text{for } W_{x-1} < L-t < W_x. \end{cases}$$

Notice that the 2 possibilities for  $x^*$  arise because one must consider whether or not the upper limit of integration is equal to a mass point of the distribution.

Alternatively, the incomplete beta probability given in (23) can be transformed to an  $F$  probability, allowing one to determine upper and lower confidence limits for  $n$  from the  $F$ -tables, using the appropriate confidence level. Applying the transformation

$$v = (n - x^* + 1) t / x^* (L - t)$$

to (23) gives

$$P(V_n \leq (n - x^* + 1) t / x^* (L - t)) = \alpha,$$

where  $V_n$  has an F distribution with parameters  $2x^*$  and  $2(n-x^*+1)$ . Notice that  $t/(L-t)$  represents the estimated odds for service of any given one of the  $n$  randomly arriving customers, as does the unobserved ratio  $(x^*/n)/\{1-(x^*-1)/n\}$ . The  $1-\alpha$  upper confidence limit for  $L$  can be determined from the equation:

$$F_{2x^*, 2(n-x^*+1)}(\alpha) = (n-x^*+1)t/x^*(L-t), \quad (24)$$

where  $F_{a,b}$  is the critical value of the central F distribution with  $a$  and  $b$  degrees of freedom, for all  $t$ ,  $0 \leq t < L$ .

The  $1-\alpha$  lower confidence limit for  $n$  can be obtained setting the conditional upper tail probability of  $T$  equal to  $\alpha$ . For the conditional upper tail probability of  $T$ ,  $x^*=x+1$  for all  $t$  (Rubin [10]). Therefore, the  $1-\alpha$  lower confidence limit for  $n$  is the solution to the equation

$$F_{2(x+1), 2(n-x)}(1-\alpha) = (n-x)t/(x+1)(L-t). \quad (25)$$

It is more convenient to apply this definition of  $x^*$  to upper and lower confidence limits, alike. Thus,  $1-2\alpha$  confidence limits for  $n$  can be constructed by holding the observed odds estimate,  $T/(L-T)$ , fixed and adjusting the unobserved odds estimate,  $\{(x+1)/n\}/\{1-(x/n)\}$ , to achieve odds ratios equal to upper and lower critical values of  $V_n$ .

Integer-valued approximate solutions to equations (24) and (25) can be determined using F-tables. Exact solutions, which are noninteger, can be computed using the F cdf and a root finding algorithm.

If one had observed  $(L, X, W_X)$ , one could use the conditional upper tail probability of  $X$ , given by (9) or (10), for deriving confidence limits of  $n$ . Since  $x \leq n$ , if  $n_{\text{lower}} < x$ , we replace  $n_{\text{lower}}$  with  $x$ . This procedure provides a  $1-2\alpha$

confidence interval which is open ( $\max(x, n_{\text{lower}}) < n < n_{\text{upper}}$ ) and has coverage probability of at least  $1-2\alpha$ . However, as mentioned earlier, this  $1-2\alpha$  confidence interval for  $n$  would be longer than that derived above using  $(L, T, w_1, \dots, w_x)$ .

The  $1-\alpha$  confidence limits for  $\lambda$  may be calculated in a similar manner, using the results of Section 3(a) when  $T$  is observed or Corollary 1.2 when  $X$  is observed. One rewrites the confidence statements based on the unconditional cdf and the unconditional upper tail probability in terms of the incomplete gamma function and solves the resulting equations for  $\lambda$ , yielding the lower and upper confidence limits, respectively.

## **6. ESTIMATION OF SERVICE TIME (w) OR RELATIVE SERVICE TIME (w/L) WHEN SERVICE TIMES ARE EQUAL**

For the case of equal service time, when both  $n$  and  $X$  are observed, one can estimate  $w$  if the shift length is known or the relative service time,  $w/L$ , if both  $w$  and  $L$  are unknown. Notice that, even for the equal service time case, one cannot resurrect  $(X, Y)$  from total idle time ( $T$ ) or total busy time ( $S$ ) when  $w$  is unknown. Maximum likelihood estimation of  $w$  based on  $(n, X, Y)$ , when  $L$  is known, is possible but is complicated by the fact that the range of  $Y$  is dependent on  $w$ . Interval estimation of  $w$  based on the joint cdf of  $X$  and  $Y$ , conditional on the number of arrivals, may be possible. We have not attempted it because the conditional joint cdf is too cumbersome. Hence, only point and interval estimation of  $w$  or  $w/L$  based on  $(n, X)$  is described below.

### 6(a) Maximum Likelihood Estimation of $w$ Based on $(n, X)$

The conditional distribution of the number of services can be used to derive the MLE of  $w$ ,  $\hat{w}_{ml}(X, n)$ . From the upper tail probability of  $X$  given in (9), notice that the conditional density function is the difference between two beta probabilities and that only the upper limit of integration for each probability depends on  $w$ . Therefore, the first derivative of the likelihood function with respect to  $w$  is proportional to

$$x^{n-x-1} (L-xw)^x - w (x-1)^{n-x+1} (L-(x-1)w)^{x-1} / (n-x). \quad (26)$$

Evaluating the second derivative of the likelihood function with respect to  $w$  at the point for which the first derivative (or, equivalently, (26) ) is equal to zero yields:

$$-x \binom{n}{x} w^{n-x} (x-1)^{n-x+1} (L-(x-1)w)^{x-1} \left\{ L/w (L-(x-1)w) + x^2 / (L-xw) \right\} < 0,$$

indicating that the MLE of  $w$  is the unique maximum of the likelihood of  $X$ , conditional on  $n$  arrivals. Using Newton's Recursion Formula with

$$w_0 = (n-x) / n(x-1)$$

as a starting value, we find that the iteration converges rapidly to  $\hat{w}_{ml}(X, n)$ .

The starting value was derived by setting equation (26) equal to zero and algebraically rearranging the equation to yield:

$$w (x-1) / (n-x) = \frac{(1-x^{-1})^{-(n-x)} (1+w/(L-xw))^{-x}}{(x/(L-(x-1)w))} \quad (27)$$

Note that the two terms in the numerator of the right hand side (RHS) of (27) are approximately equal to one and that the denominator of the RHS is greater than one for  $x=1, \dots, n$ , and is approximately equal to  $n$ , for  $x=n$  and small  $w$ . Thus,

$$1/w \cong n(x-1)/(n-x).$$

### 6(b) Confidence Limits for w Based on (n, X)

The conditional upper tail probability of the number of services is used to derive confidence limits for w when (n, X) are observed and the shift length is known. Because n and w are inversely related, we use the equation that yields the upper confidence limit for n to yield the lower confidence limit for w, and proceed similarly for the upper confidence limit of w. Solving for w in

$$\frac{n-x}{x+1} \left( \frac{L-xw}{xw} \right) = F_{2(x+1), 2(n-x)}(1-\alpha) \quad (28)$$

yields the  $1-\alpha$  lower confidence limit while solving equation for w in

$$\frac{n-x+1}{x} \left( \frac{L-(x-1)w}{(x-1)w} \right) = F_{2x, 2(n-x+1)}(\alpha) \quad (29)$$

yields the  $1-\alpha$  upper confidence limit. These equations can be solved explicitly for w, while they must be solved iteratively to yield confidence limits for n.

Therefore, the  $1-2\alpha$  confidence interval for w based on (n, X) is:

$$\frac{L(n-X)}{X((n-X)+(X+1)F_{2(x+1), 2(n-x)}(1-\alpha))} < w < \frac{L(n-X+1)F_{2(n-x+1), 2x}(1-\alpha)}{(X-1)(X+(n-X+1)F_{2(n-x+1), 2x}(1-\alpha))}.$$

The  $1-2\alpha$  confidence interval for w/L based on (n, X) is found by dividing through by L in the expression given above.

When equation (29), or its generalized version for arbitrary service distribution, is used to calculate a lower bound on n, if  $n_{\text{lower}}(x) < x$ , we replace  $n_{\text{lower}}(x)$  with x. Likewise, we replace the upper confidence limit of w with L if the upper confidence limit exceeds L. Therefore, when  $X=1$  both the MLE and the upper confidence limit of w are equal to L for all n, but the lower confidence limit of w depends on n. As n approaches infinity, the lower confidence limit of w

approaches the limit  $L/X$  for all values of  $X > 1$ , while the upper confidence limit approaches  $L/(X-1)$  for  $X > 1$ .

## APPENDIX A

Corollary 2.4. When all services are of duration  $w$  and  $L=1$ , the probability of no overtime, conditional on the number of arrivals, is:

$$P(Y_n = 0) = 1 - nw P(Y_{n-1} = 0) \quad \text{for } n < [1/w].$$

Proof. Using the pdf of  $Y_n$  for the case  $n < [1/w]$  yields:

$$P(Y_n = 0) = \sum_{x=1}^n \binom{n}{x} (1-xw)^x (xw)^{n-x} \quad \text{for } nw < 1.$$

The recursion relation is proved by showing that

$$P(Y_n = 0) + nw P(Y_{n-1} = 0) = 1$$

if  $0 < nw < 1$ , which is equivalent to showing that

$$-1 + \sum_{x=1}^n \binom{n}{x} (1-xw)^x (xw)^{n-x} + nw \sum_{x=1}^{n-1} \binom{n-1}{x} (1-xw)^x (xw)^{n-1-x} = 0. \quad (\text{A.1})$$

Noting that

$$nw \sum_{x=1}^{n-1} \binom{n-1}{x} (1-xw)^x (xw)^{n-1-x} = \sum_{x=1}^n \binom{n}{x} (1-xw)^x (xw)^{n-x} (n-x) / x,$$

we find that the left hand side (LHS) of (A.1) is equal to

$$-1 + (1-nw)^n + nw \sum_{x=1}^{n-1} \binom{n}{x} (1-xw)^x (xw)^{n-x-1}.$$

Applying the Binomial Theorem to  $(1-nw)^n$  and  $(1-xw)^x$  and combining these terms yields:

$$\text{LHS} = nw \left\{ \sum_{x=1}^n \binom{n}{x} (-1)^x (nw)^{x-1} + \sum_{x=1}^{n-1} \binom{n}{x} \sum_{v=0}^x \binom{x}{v} (-1)^v (xw)^{n-x-1+v} \right\}.$$

Making the transformation  $k=x-v$  and noting that

$$\binom{n}{x} \binom{x}{x-k} = \binom{n}{k} \binom{n-k}{x-k}$$

give:

$$\text{LHS} = nw \sum_{k=0}^{n-1} \binom{n}{k} (w)^{n-1-k} \sum_{x=k}^n \binom{n-k}{x-k} (-1)^{x-k} (x)^{n-k-1}.$$

Making the transformation  $t = x-k$  gives:

$$\text{LHS} = nw \sum_{k=0}^{n-1} \binom{n}{k} (w)^{n-1-k} \sum_{t=0}^{n-k} \binom{n-k}{t} (-1)^t (t+k)^{n-k-1}.$$

Letting  $m=n-k-1$  and applying the Binomial Theorem to  $(t+k)^{n-k-1} = m$  give:

$$\text{LHS} = nw \sum_{k=0}^{n-1} \binom{n}{k} (w)^{n-1-k} \sum_{t=0}^{m+1} \binom{m+1}{t} (-1)^t \sum_{v=0}^m \binom{m}{v} k^v t^{m-v}.$$

Switching the order of summation for  $t$  and  $v$  gives:

$$\text{LHS} = nw \sum_{k=0}^{n-1} \binom{n}{k} (w)^{n-1-k} \sum_{v=0}^m \binom{m}{v} k^v \sum_{t=0}^{m+1} \binom{m+1}{t} (-1)^t t^{m-v}. \quad (\text{A.2})$$

Note that

$$\sum_{t=0}^{m+1} \binom{m+1}{t} (-1)^t t^{m-v} = 0 \quad \text{for } v = 0, 1, 2, \dots, m$$

(Gradshteyn and Ryzhik [5], p. 4). Since (A.2) is a polynomial in  $k$  with the coefficient of each term equal to zero, then (A.2) is equal to zero. Therefore, (A.1) is equal to zero and

$$P(Y_n = 0) = 1 - nw P(Y_{n-1} = 0).$$

□

## APPENDIX B

Corollary 2.6. When all services are of duration  $w$  and  $L=1$ , the probability that overtime exceeds  $y$ , conditional on the number of arrivals, is:

$$P(Y_n > y) = n(w - y) P(Y_{n-1} = 0)$$

for  $n < [1/w]$  and  $0 < y < w < 1$ .

Proof. Let  $D = w - y$ .

(Note that  $P(Y_n > y) = P(w - Y_n < w - y \equiv D)$ .)

$$\begin{aligned} P(Y_N > y) &= \sum_{x=0}^{n-1} P(Y_N > y, X_N = x+1) \\ &= P(Y_N > y, X_N = 1) + \sum_{x=1}^{n-1} P(Y_N < y, X_N = x+1). \end{aligned} \quad (B.1)$$

For  $n$  arrivals and only one service, overtime exceeds  $y$  only when all  $n$  arrivals occur in the interval  $(1-w+y, 1)$ . Consequently,  $P(Y_n > y, X_n = 1) = (w-y)^n = \Delta^n$ .

The second term in (B.1) is most easily calculated in instant service mode as

$$\sum_{x=1}^{n-1} \sum_{a, b, c} (n! / (a! b! c!)) (1-xw-\Delta)^a \Delta^b (xw)^c,$$

where  $a = 0, 1, \dots, x$ ,  $b = 1, 2, \dots, n$  and  $c = n - (a + b)$ . With  $r \equiv a + b$ , applying the Binomial Theorem to  $(1-xw-\Delta)^a$  gives:

$$\sum_{x=1}^{n-1} \sum_{r=x+1}^n \binom{n}{r} (xw)^{n-r} \sum_{b=n-x}^r \binom{r}{b} \sum_{v=0}^{r-b} \binom{r-b}{v} (-1)^v \Delta^{b+v} (1-xw)^{r-(b+v)}.$$

With  $k \equiv b + v$ , we must calculate the coefficients of  $\Delta^k$  for the polynomial in the equation above. Applying the procedure developed in the proof of Corollary 2.4, we find that the coefficient of  $\Delta^n$  in (B.1) is:

$$\sum_{a=0}^{n-1} \binom{n}{a} (-1)^a (n-a) = 0$$

and that for  $\Delta^k$  with  $1 < k < n$  the coefficient is:

$$\sum_{x=1}^{n-1} \sum_{r=x+1}^{x+k} \binom{n}{r} (xw)^{n-r} (1-xw)^{n-k} \sum_{b=n-x}^k \binom{r}{b} \binom{r-b}{k-b} (-1)^{k-b}. \quad (B.2)$$



After several transformations we find that (B.2) equals

$$\binom{n}{k} \sum_{g=0}^{n-k} ((n-k)!/g!) (-w)^{n-k-g} \sum_{b=0}^{k-1} \binom{k-1}{b} (-1)^{k-b},$$

where

$$\sum_{b=0}^{k-1} \binom{k-1}{b} (-1)^{k-b} = 0.$$

Therefore, the coefficient of  $\Delta^k$  is zero for  $1 < k < n$ . Since (B.2) holds for  $k=1$ , the coefficient for  $\Delta$  is:

$$\begin{aligned} \sum_{x=1}^{n-1} \binom{n}{x+1} (xw)^{n-x-1} (1-xw)^x \binom{x+1}{1} &= n \sum_{x=1}^{n-1} \binom{n-1}{x} (1-xw)^x (xw)^{n-1-x} \\ &= n P(Y_{n-1}=0). \end{aligned}$$

Therefore,

$$P(Y_n > y) = n (w - y) P(Y_{n-1} = 0). \quad \square$$

## REFERENCES

- [1] I. V. Basawa and B. L. S. Prakasa Rao, Statistical Inference for Stochastic Processes, Academic Press, Inc., London, 1980.
- [2] I. V. Basawa and N. U. Prabhu, Large Sample Inference from Single Server Queues, Queueing Systems 3 (1985) 289-304.
- [3] D.R. Cox and W. L. Smith, Queues, Methuen and Co., Ltd., London, 1961.
- [4] W. Feller, An Introduction to Probability Theory and Its Applications (third edition), Volume 1, John Wiley and Sons, New York, 1968.
- [5] I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products (Corrected and Enlarged Edition), Academic Press, Inc., New York, 1980.

- [6] N. L. Johnson and S. Kotz, Distributions in Statistics: Discrete Distributions, John Wiley and Sons, New York, 1969.
- [7] E. L. Lehmann, Testing Statistical Hypotheses (second edition), John Wiley and Sons, New York, 1986.
- [8] N. U. Prabhu, Stochastic Storage Processes: Queues, Insurance Risk, and Dams, Springer- Verlag, New York, 1980.
- [9] D. S. Robson and C. Jones, The Theoretical Basis of an Access Site Angler Survey Design, *Biometrics* 45 (1989) 83-98.
- [10] G. Rubin, Statistical Distribution and Estimation Theory for a Single Server Queue with Fixed Service Time and Complete Balking. Ph.D. Dissertation, Cornell University, Ithaca, NY, 1987.
- [11] G. Rubin and D. S. Robson, Estimation Theory for a Single Server Queue with Random Arrivals and Complete Balking. Technical Report #BU-1020-M in Biometric Series, Cornell University, Ithaca, NY, 1989.
- [12] J. E. Samaan and D.S. Tracy, On the conditional estimation for a parameter of a queueing system with loss. In: Computer Science and Statistics: 12<sup>th</sup> Annual Symposium on the Interface, J. F. Gentleman, Ed. (The University of Waterloo, Waterloo, Ontario, 1979).

## LIST OF FIGURES

Figure 1. Cdfs of total server idle time (T) with  $w=0.1$  and  $L=1$ : unconditional ( $\Delta$ ) with  $\lambda=20$  and conditional (\*) with  $n=20$ .

Figure 2. Expected number of services, conditional on  $n=15$ , vs.  $w$  for  $L=10$  ( $\Delta$  = exact; \* = approximate).